

Inherited Platelet Disorders: Towards DNA-based diagnosis

Claire Lentaigne^{1,2}, Kathleen Freson³, Michael A Laffan^{1,2}, Ernest Turro⁴⁻⁷, Willem H Ouwehand^{4,5,7,8~}

¹ Centre for Haematology, Hammersmith Campus, Imperial College Academic Health Sciences Centre, Imperial College London, London, United Kingdom.

² Imperial College Healthcare NHS Trust, Du Cane Road, London, United Kingdom.

³ Department of Cardiovascular Sciences, Center for Molecular and Vascular Biology, University of Leuven, Belgium.

⁴ Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, United Kingdom.

⁵ NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, United Kingdom.

⁶ Medical Research Council Biostatistics Unit, Cambridge Institute of Public Health, Cambridge Biomedical Campus, Cambridge, United Kingdom.

⁷ NIHR BioResource - Rare Diseases, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, United Kingdom.

⁸ Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom.

~on behalf of the BRIDGE Bleeding, Thrombotic and Platelet Disorders and ThromboGenomics Consortia.

Abstract

Variations in platelet number, volume and function are largely genetically controlled and many loci associated with platelet traits have been identified by genome wide association studies (GWAS)¹. The genome also contains a large number of rare variants, of which a tiny fraction underlie the inherited diseases of man. Research over the past three decades have led to the discovery of 51 genes harbouring variants responsible for inherited platelet disorders (IPDs). However, the majority of patients with an IPD still do not receive a molecular diagnosis. Alongside the scientific interest, molecular or genetic diagnosis is important for patients. There is increasing recognition that a number of IPDs are associated with severe pathologies, including an increased risk of malignancy and a definitive diagnosis can inform prognosis and care.

In this review we give an overview of these disorders grouped according to their effect on platelet biology and their clinical characteristics. We also discuss the challenge of identifying candidate genes and causal variants therein, how IPDs have been historically diagnosed and how this is changing with the introduction of high-throughput sequencing (HTS). Finally, we describe how integration of large genomic, epigenomic and phenotypic datasets, including whole genome sequencing (WGS) data, GWAS, epigenomic profiling, protein-protein interaction networks and standardised clinical phenotype coding, will drive the discovery of novel mechanisms of disease in the near future to improve patient diagnosis and management.

Rare Inherited Platelet Disorders

There is marked genetic heterogeneity amongst IPDs and in this section we survey the 51 genes known to harbour variants responsible for IPDs (henceforth, “IPD genes”), classified according to their principal known effect on platelet biology. They encode an array of molecules of diverse function, reflecting the complex and tightly regulated processes of megakaryopoiesis, platelet formation and platelet function (**Fig. 1**). For some genes, their role in platelet biology is less well defined and this is also discussed. Many IPD genes are widely transcribed across blood cell types (**Fig. 2**) and other tissues. Hence, patients with an IPD frequently present with pathologies reaching well beyond the blood system² (23 genes marked with *, **Fig. 1**).

Megakaryopoiesis and Platelet Formation

Studies in patients with thrombocytopenia have highlighted the important roles of the Thpo/Mpl signalling pathway and transcriptional regulation in both early and late stages of megakaryopoiesis. Rare variants in *THPO* and *MPL*, the genes encoding thrombopoietin^{3,4} and its receptor Mpl^{5,6} cause congenital thrombocytopenia and seven IPD genes encode transcription factors (*GATA1*^{7,8}, *RUNX1*⁹, *FLI1*¹⁰, *ETV6*¹¹, *HOXA11*¹², *MECOM*¹³ and *GFI1B*^{14,15}) expressed in hematopoietic stem and progenitor cells (**Fig. 2**). Rare variants in *RBM8A*¹⁶ and *ANKRD26*¹⁷ have been shown to affect transcription factor binding (of Mecom and Runx1, respectively), altering signalling through the Thpo/Mpl pathway and resulting in thrombocytopenia, but their exact role in megakaryopoiesis is not yet clear.

Six IPD genes regulate the megakaryocyte cytoskeleton and the principal effect of deleterious variants is macrothrombocytopenia due to aberrant proplatelet formation (variants in *MYH9*¹⁸, *ACTN1*¹⁹, *FLNA*²⁰, *TUBB1*²¹, *DIAPH1*²²). Causal variants in the Wiskott-Aldrich Syndrome gene *WAS* also lead to defective proplatelet formation alongside neutropenia and eczema but, in contrast to other cytoskeletal disorders, the platelets are small^{23,24}. So far six genes encoding proteins that primarily influence granule formation, trafficking or secretion (*NBEAL2*²⁵⁻²⁷, *NBEA*²⁸, *VPS33B*²⁹, *VIPAS39*³⁰, *STXBP2*^{31,32}, *LYST*³³) and nine genes that cause Hermansky Pudlak Syndrome (HPS), a δ-

granule platelet disorder (*HPS1*, *AP3B1*, *HPS3*, *HPS4*, *HPS5*, *HPS6*, *DTNBP1*, *BLOC1S3*, *BLOC1S6*³⁴⁻⁴¹) have been identified.

Defects in transmembrane glycoprotein (GP) signalling pathways can lead to abnormal platelet function and thrombocytopenia with giant platelets through abnormal proplatelet formation (*GP1BA*, *GP1BB*, *GP9*^{42,43}) and thrombasthenia only (*ITGA2B*, *ITGB3*^{44,45}). Gain-of-function (GOF) variants in *GP1BA* can cause enhanced binding to Von Willebrand Factor (Vwf) leading to “platelet-type” Von Willebrands Disease (VWD)⁴⁶, a phenocopy of Type 2B VWD, a disorder caused by GOF variants in *VWF* affecting the function of its A1 domain⁴⁷. In both cases premature interaction between Gp1ba and the Vwf A1 domain results in bleeding characterised by thrombocytopenia and a loss of high molecular weight Vwf multimers. DNA analysis is generally required to distinguish between these two disorders.

The role of some IPD genes in megakaryopoiesis and proplatelet formation is less well defined. The recent discovery of a GOF variant in *SRC* causing abnormal megakaryopoiesis and thrombocytopenia has cemented the central role of this universal tyrosine kinase in a variety of megakaryocyte signalling pathways and podosome formation⁴⁸; *CYCS* is expected to play a role in apoptosis⁴⁹; certain GOF variants in *STIM1* (causal of Störmorken syndrome) result in thrombocytopenia with abnormal platelet function and bleeding, whereas loss-of-function (LOF) variants in *STIM1* cause an autoimmune thrombocytopenia. LOF variants in *ORAI1* have been described in a mild Störmorken-like syndrome but the platelet defects are not well established. Both *Orai1* and *Stim1* are involved in calcium homeostasis and IPD-causing variants consequently affect platelet signalling but further research is required into their role in platelet formation⁵⁰⁻⁵². Novel missense variants were recently identified in *SLFN14* in four unrelated pedigrees with moderate thrombocytopenia and platelet secretion defects. Though the mechanism is not clear, a defect in platelet formation was observed^{53,54}. Finally, Quebec platelet disorder, so far confined to French-Canadians, is caused by a tandem duplication in *PLAU*, leading to overexpression of urokinase and is characterised by thrombocytopenia, degradation of α -granule contents with normal granule structure, decreased aggregation in response to epinephrine and late onset bleeding⁵⁵.

Platelet Function

There are 14 IPD genes primarily affecting various aspects of platelet function. Six genes encode GPs which function as receptors for the hemostatically important ligands Vwf (*GP1BA*, *GP1BB*, *GP9* gene defects causing Bernard Soulier syndrome {BSS})⁴³, fibrinogen (*ITGA2*, *ITGB3* gene defects causing Glanzmann Thrombasthenia {GT})⁴⁵ and collagen (*GP6*)⁵⁶. The typical mode of inheritance for GT is autosomal recessive with LOF variants of *ITGA2B* or *ITGB3*. This classical thrombasthenia is reviewed extensively elsewhere⁴⁵, but it is worth noting that, again, rare and dominant GOF variants in *ITGB3* have also been described, leading to enhanced fibrinogen binding combined with bleeding⁵⁷ and a mild thrombocytopenia exacerbated during pregnancy but without bleeding⁵⁸. Mutations in *FERMT3* affect integrin inside-out signalling causing a GT-like platelet phenotype with bleeding and associated with a type III leukocyte adhesion disorder⁵⁹. A homozygous variant in *RASGRP2* was discovered as a cause of another mild GT-like platelet phenotype in a consanguineous pedigree. *RASGRP2* encodes for a guanine exchange factor that regulates Rap1b activation and consequently has a major effect on $\alpha\text{IIb}\beta 3$ (encoded by *ITGA2B*, *ITGB3*) signalling in platelets⁶⁰.

G-protein coupled receptors (GPCRs) are another main type of multispan transmembrane receptors and signalling defects due to variants in *P2YR12*⁶¹ and *TBXA2R*⁶² (encoding the GPCRs for ADP and thromboxane respectively) or their downstream effectors (*TBXAS1*⁶³, *PLA2G4A*⁶⁴) have been linked to IPDs. Scott syndrome does not fall into the aforementioned categories and is caused by autosomal recessive variants in *ANO6*, which encodes a multispan transmembrane protein involved in phospholipid scrambling⁶⁵. Platelets from these cases cannot properly express phosphatidylserine on their surface, which leads to defective coagulation.

Current Approaches to Diagnosis

IPD diagnosis is straightforward in the major platelet function disorders such as BSS and GT, which often present with severe bleeding symptoms early in life and are easily recognised by the pattern of platelet aggregation defects^{45,66}. This is often supplemented by assessment of the storage pool either directly by nucleotide assay or

lumi-aggregometry. In some IPDs, the platelet function defect and impaired hemostasis are part of well-defined syndromes, e.g. Chediak Higashi Syndrome (CHS) and HPS, where the platelet δ -granule defect is typically associated with immune deficiency or ocular albinism, respectively. The presence of syndromic features can help in recognition and diagnosis but diagnosis remains challenging for the majority of IPDs, which often have a mild platelet phenotype and are clinically heterogeneous. Diagnosis is further complicated by the fact that for many IPDs, the platelet count is within normal ranges and the disorder may only become apparent after a hemostatic challenge or if cases present with accompanying pathologies in other organ systems, including malignancies⁶⁷.

Establishing a conclusive molecular diagnosis is the bedrock of good hematological practice because it informs optimal treatment and can provide clarity about disease progression. For IPDs, this is particularly important for the severe cases and those associated with early onset clinical pathologies such as myelofibrosis, lung fibrosis, renal insufficiency and malignancy. Thrombocytopenias caused by variants in *RUNX1*, *ETV6* and *ANKRD26*^{9,68,69} are associated with increased risk of myeloid malignancy while for WAS and amegakaryocytic thrombocytopenia caused by *MPL* variants, treatment by allogeneic hematopoietic stem cell transplant or gene therapy may require consideration. Moreover, genetic counselling can be provided if the diagnosis is confirmed at the DNA level⁷⁰⁻⁷². Current guidelines favour a tiered approach to IPD diagnosis. DNA analysis by Sanger sequencing is at the fourth and final tier and often not applied because of its limited availability and costs⁶⁶. Moreover, it is used primarily to confirm an *already clinically suspected genetic diagnosis* and targets only a single or small group of genes. In the majority of IPDs, a single candidate gene is not readily apparent from standard laboratory tests. Consequently, a molecular diagnosis is given in only a minority of patients and even when a genetic defect is identified, the number of independent cases remains small for the majority of IPDs^{73,74}. Fewer than five unrelated probands have been identified for IPD genes *P2Y12R*⁷⁵, *GP6*^{76,77}, *TBXA2R*⁷⁸, *PLAU*, *ANO6*⁷⁹, which all were identified before the era of HTS and there is a paucity of larger case series for individual IPD genes with the exception of *ACTN1*⁸⁰ and *ANKRD26*⁸¹.

This lack of genetic diagnosis not only hampers our ability to provide accurate information on prognosis and optimal management, but the small case numbers also impact on our ability to interpret the pathogenicity of particular variants. The advent of HTS is set to change this. In this issue of *Blood*, the ThromboGenomics consortium reports on a targeted HTS panel of 76 genes (63 genes in the panel reported in their publication and a further 13 genes added in the currently available version) covering the inherited bleeding, thrombotic and platelet disorders (BPDs), bringing an affordable molecular diagnosis within reach⁸². The application of HTS will simplify the diagnostic process and reduce delay, as we discuss below.

IPD Gene Discovery to Date

Until recently, the majority of IPD genes have been discovered by candidate gene and linkage studies. These approaches resulted in the identification of 36 IPD genes by 2010, when HTS became available (**Fig. 3**). In 2011 three groups reported on *NBEAL2* being the causal gene for Gray Platelet syndrome (GPS)²⁵⁻²⁷ and two of these discoveries were made possible by HTS. In the past five years 14 additional IPD genes have been identified, 11 by applying HTS (**Fig. 3**). Thrombocytopenia with absent radii (TAR) is an example of a syndrome for which the genetic roots remained elusive despite being clinically well defined for over two decades. It has a thus far unique genetic architecture typically involving a micro-deletion on one haplotype and a low-frequency regulatory variant on the other haplotype of *RBM8A*. The most commonly implicated regulatory variant results in reduced binding of the transcription factor Mecom leading to an insufficiency of the protein Y14 in megakaryocytes, causing thrombocytopenia¹⁶. In addition, WES was also instrumental in discovering variants in *GFI1B* responsible for a GPS-like syndrome^{14,15}; *ACTN1*¹⁹, *ETV6*¹¹, *STIM1*⁵⁰, *DIAPH1*²², *SRC*⁴⁸, *SLFN14*⁵³ and *MECOM*¹³ for inherited thrombocytopenias; and *RASGRP2* for GT-like disease⁶⁰.

The results of the rare diseases pilot phase of the 100,000 Genomes Project⁸³ indicate that a large number of IPD genes remain to be discovered. Genome sequencing results of the DNA samples of hundreds of probands with uncharacterised BPDs, analysed using assigned Human Phenotype Ontology (HPO) terms² have helped identify

pathogenic variants in known IPD genes in almost 20% of cases. New clustering algorithms to group cases with similar phenotypes^{48,84}, have been used to identify two novel IPD genes (*DIAPH1* and *SRC*)^{22,48} and several putative ones. This suggests that the majority of cases either harbour pathogenic variants in unknown genes or regulatory regions or are the result of a digenic mode of inheritance.

The lack of gene discoveries on a global scale for δ -granule Storage Pool Disease (δ -SPD) is a case in point. The true prevalence of δ -SPD is unclear, as the definition is not always consistent between studies, but defects in granule release are relatively frequent amongst the mild IPDs^{73,85}. The greatest diagnostic success for δ -SPD has been in the multi-system syndromes such as CHS and HPS but most patients with non-syndromic δ -SPD remain undiagnosed⁸⁶. In on-going large-scale HTS projects such as GAPP^{87,88} and BRIDGE-BPD^{2,22,48}, the lack of gene discovery for δ -SPD, even in these large patient cohorts, highlights a need to analyse the non-coding regulatory regions of the genome via WGS, whilst also exploring novel methods of data analysis and integration. We discuss such methods in more detail in the final section.

Assigning Pathogenicity to Novel Variants: Use of Public Databases

With the advent of HTS it has become possible to survey the exonic fraction of large numbers of genomes for variants by WES. Exomes comprise ~2% of the genome (~64 Mb) while current WGS captures 98.3% of the 3.2 billion bases of the genome at a minimum of 15x coverage. Information about pathogenic and likely pathogenic variants is maintained in databases but until recently it was not possible to verify their allele frequencies in the general population. Several initiatives such as the 1000 Genomes⁸⁹ and UK10K⁹⁰ projects and initiatives to aggregate results from many smaller WES projects, as achieved by the Exome Aggregation Consortium (ExAC)⁹¹, have provided information on exonic variants observed in more than 71,000 individuals of mainly Caucasoid ancestry. This catalogue has made an immense contribution to improving the accuracy of assignment of pathogenicity to DNA variants observed in IPD cases. There is still a relative paucity of sequencing data from individuals of other ethnicities. Consequently, a larger number of variants absent from control samples is

observed in their DNA, making it harder to distinguish variants which are relatively common in particular populations from variants responsible for disease.

HTS has also made possible large-scale WGS projects such as the 100,000 Genomes Project⁸³ which are complemented by GWAS in large population cohorts such as the UK Biobank of 500,000 healthy individuals⁹². Both projects link genotypes with health and social care records and will lead to a better understanding of the relationship between variants and diseases. In particular, they will allow more accurate assignment of pathogenicity to rare variants underlying the thousands of inherited diseases. Single nucleotide variants (SNV) can be called accurately by WES but calling of short insertions and deletions and especially of copy number and structural variants is more challenging. Here WGS can achieve far greater sensitivity and specificity than WES. WGS will therefore further improve the accuracy of frequencies for all classes of variants in databases, make it easier to discover the genetic determinants of rare inherited diseases and also open up opportunities to explore the non-coding regulatory part of the genome.

Germline variants are inevitable consequences of meiosis and DNA repair, and accumulate over generations. Given the existence of many non-pathogenic variants in any individual's genome, the main challenge faced by researchers when interpreting HTS data of an IPD case is determining which variants are causing the disorder. The correct identification of novel causal variants critically depends on their allele frequency in relevant control samples⁹³. Additionally, studies are required to uncover the function of novel genes and the consequences of candidate rare variants. It has been shown that specific GOF variants in *DIAPH1* and *SRC* can cause a defect in megakaryopoiesis while this is not expected of LOF variants^{22,48}. Similar observations can be made for *SFLN14* where all variants are located in the ATPase-AAA-4 domain while rare variants outside this domain seem not to result in an IPD⁵⁴.

There are publicly accessible databases like ClinVar⁹⁴ and DECIPHER⁹⁵, Exome variant server⁹⁶ and access-for-a-fee databases like Human Gene Mutation Database (HGMD)⁹⁷ that record disease-associated variants. HGMD maintains a catalogue of high-penetrance variants derived from the literature⁹⁷. None of these databases are 100% accurate: For example, 539 rare variants denoted disease-causing in HGMD were

observed in the 1000 Genomes Project at a frequency above 1%⁹⁸ and 140 variants labelled as causing BPDs showed a frequency in ExAC above 0.1%, yet the evidence supporting a claim to pathogenicity was deemed insufficient for all but four variants⁸².

In conclusion, large-scale population sequencing projects have greatly enhanced our ability to interpret the pathogenicity of potential candidate IPD-causing variants, but there are pitfalls: any rare variant must be interpreted in the context of the ethnicity of the individual and great care must be taken not to over-interpret the pathogenicity of novel variants absent from control datasets, even in established IPD genes, without back-up from functional studies or observation of the same variant in several unrelated cases with similar phenotypes.

The Future of IPD Gene Discovery

HTS will undoubtedly assist the discovery of variants causing IPDs in many new genes over the next decade but genomic sequencing alone cannot explain the mechanisms underlying the relationship between genotype and phenotype in cases with an apparent inherited disorder⁹⁹. In order to ascertain the functional consequences of rare variants, it is essential that knowledge of phenotypes and pathways is integrated systematically within a frame of reference similar to that of the human genome¹⁰⁰.

Data from a wide variety of sources, including GWAS, Online Mendelian Inheritance in Man (OMIM) and mouse genome databases can be used to annotate candidate regions, better understand individual proteins and their roles in pathways relevant to megakaryopoiesis and platelet formation. Identification of the regulatory DNA elements has become feasible thanks to the results from projects like ENCODE¹⁰¹, Roadmap¹⁰² and Blueprint¹⁰³, coordinated by the International Human Epigenome Consortium (IHEC). Integration of different layers of information, including methylation and histone modification states, expression quantitative trait loci data, transcriptomics and proteomics, requires the development of new statistical methods. Some insights on how this richness of information can aid the discovery of novel causes of disease will be discussed in the following sections.

Human Phenotype Ontology

To maintain power of gene discovery, cases sharing similar clinical and laboratory phenotypes need to be identified and new methods have been sought to cluster similar cases in the ever-growing cohorts undergoing genome sequencing. One of the widely used phenotype annotation standard for rare diseases is the HPO terms system¹⁰⁴. HPO coding is used by the Deciphering of Developmental Disorders¹⁰⁵ and 100,000 Genomes Projects and so far 1,247 probands with BPDs have been HPO coded². This revealed the presence of non-hematological pathologies in 60% of cases, particularly in the central nervous (e.g. autism spectrum disorder), skeletal (e.g. osteoporosis) and immune systems². This insight into the more complex spectrum of pathologies in IPD cases is important for the provision of care, which often warrants a multi-disciplinary approach. Additionally, standardised phenotyping by means of HPO terms is critical for IPD gene discovery across large collections of cases. Indeed, genome sequencing combined with HPO coding supported the identification of the *DIAPH1* variant in two unrelated pedigrees with similar phenotype terms 'Thrombocytopenia' and 'Deafness'²². It also allowed integration with existing phenotype databases, such as the one for mouse phenotype ontology (MPO) terms, which aided the discovery of the *SRC* variant because cases and knockout mice shared HPO terms⁴⁸. These discoveries are critically dependent on new statistical methods which exploit the power of HPO-based patient coding together with genotypes obtained by sequencing^{84,106,107}.

In order to discover which genes are pertinent to the remaining IPDs, screening of large case collections will be essential. The small number of reported independent cases in the majority of IPDs and the lack of discovery in SPD to date indicates that extremely large collections are needed to bring together adequate numbers of unrelated index cases with a shared genetic basis. International collaboration is therefore warranted and will also bring together expertise about the clinical evolution of disease for each of the IPD genes. This collaborative approach will provide the platform to evaluate existing and new interventions to gather evidence for the best approach to treatment. That such international approaches can be successful has been demonstrated by the BRIDGE-BPD and ThromboGenomics consortium efforts reported in this issue^{22,82} and in other journals^{48,84}.

The Non-coding Regulatory Space

The genome comprises 3.2 billion base pairs, 98% of which do not form part of a recognised gene. The non-exonic portion of the genome is largely regulatory in nature but the landscape of regulatory elements differs between cell types. The aforementioned IHEC initiatives have generated accurate maps of regulatory elements. The discoveries that specific heterozygous GOF variants in the 5'UTR of *ANKRD26* cause thrombocytopenia¹⁰⁸ and that the vast majority of TAR syndrome cases of Northern European ancestry are due to compound inheritance of one *RBM8A* null allele and a low-frequency SNV in its 5'UTR¹⁶ indicate that cell-specific knowledge of the regulatory space may help to identify novel mechanisms of disease in future. It is likely that similar non-coding regulatory variants in known or novel IPD genes remain to be discovered and integrating IHEC reference epigenome maps with the catalogue of blood cell-type-specific transcript isoform usage can help interpret the consequence of non-coding DNA variants observed in IPD cases¹⁰³. The exploration of the non-coding portion of the genome for variants causal of IPDs requires many parallel approaches recognising the complexity of the regulatory networks at play. For example the cell-specific role of non-coding RNAs has been recently highlighted¹⁰⁹. Finally the rapidly accumulating ChIP-seq data defining the binding sites for transcription factors in different cell types, including data generated in megakaryocytes will also be of value¹¹⁰.

GWAS for Platelet Traits

Meta-analysis of GWAS for blood cell traits has led to the discovery of nearly 200 common SNVs exerting small effects on blood cell indices¹¹¹. These variants mark known and new genomic regions important for hematopoiesis and blood cell survival. With GWAS being performed on ever-larger population samples such as the UK Biobank and Million Veteran cohorts, the number of associated SNVs is predicted to rise substantially, with increasing numbers of rare variants with larger effect sizes being revealed. A picture is therefore emerging of hundreds of genes identified by GWAS controlling the life cycle of platelets and the maximum effect sizes of variants in these genes on platelet traits are inversely correlated with their MAFs due to selection. This assumption is illustrated by recent observations in *TUBB1*. The common non-

coding SNV rs4812048 was linked with an effect on platelet volume by GWAS¹ and rare non-synonymous SNVs have since been reported as a cause of macrothrombocytopenia in man^{21,112-114}. This suggests that integrating knowledge about GWAS loci affecting platelet indices alongside genome sequencing data of collections of patients with IPDs of unknown molecular etiology may reveal novel candidate genes.

Protein-protein Networks

Knowledge about the molecules and pathways underlying megakaryopoiesis and the formation of platelets is rapidly expanding. Bringing knowledge gathered from GWAS¹¹¹ and inherited BPDs (the 76 BPD genes reported by Simeoni *et al*, this issue⁸²) together with information about genes identified by ChIP-seq¹¹⁰ we defined 200 unique genes deemed important for these processes (**Fig. 4B**). The proteins encoded by these “seed genes” were used as baits to retrieve their first-order interactors from the Reactome and IntAct databases using previously reported informatics approaches^{1,115}. This information was displayed as a protein-protein interaction network (PPIN) consisting of 1,684 nodes (proteins) connected by 5,360 biochemical interactions (edges) using the Cytoscape application (**Fig. 4A**). To make this PPIN, which encompasses the knowledge from thousands of publications available to the scientific and medical communities in a format allowing its unrestricted use, we have made a Cytoscape file available for download (Supplemental data). For example, it can be used to resolve IPDs with digenic roots as illustrated by observations in *TBXA2R*: four reports describe LOF variants of *TBXA2R* in patients with platelet defects as a dominant finding. Although heterozygosity for the *TBXA2R* variants correlated with the platelet defect in these carriers, there was no association with bleeding problems. It was noted that a potential second genetic factor would be required to cause bleeding. Two sub-networks (**Fig. 4C, D**) of proteins important for the signalling events downstream of the thromboxane receptor (**Fig. 4C**) and for the synthesis of thromboxane (**Fig. 4D**) may highlight potential candidates for this second gene^{78,116-118}. Indeed we observed a case with severe bleeding and a LOF variant in *TBXA2R* (**Fig. 4C**) and a second putative causal variant was identified in *PTGS1*, which encodes for Cyclooxygenase-1¹¹⁹.

Conclusion

HTS has enabled the discovery of many novel IPD genes in the past five years and this new knowledge can be rapidly integrated in diagnostic platforms like the ThromboGenomics HTS test, which will simplify and hasten diagnosis of IPDs. There is however still an immense challenge to resolve the genetic basis of the remaining IPDs and to gather better evidence for the best treatments. This can only happen through international collaboration and knowledge sharing. We must seek permission from patients and their families to share their genotype and phenotype data and invest in standardized phenotyping using internationally agreed terms like those of the HPO system. There is also an obligation for the research and clinical communities to pursue the development of informatics environments for the safe sharing of anonymized and linked-anonymized data. WGS combined with emerging data from GWAS, ChIP-seq, proteomics and mouse knock-out studies amongst others will also help explore the non-coding regulatory space and identify novel candidate IPD genes and variants.

Finally, as there are many potential pitfalls when interpreting the role of novel rare variants, it is important to apply rigorous standards when assigning pathogenicity. Providing a molecular diagnosis to patients is highly desirable but making incorrect assumptions about variants could be harmful.

Acknowledgments

We would like to thank the members of the BRIDGE-BPD and ThromboGenomics Consortia for their contributions. The BRIDGE-BPD and ThromboGenomics studies, including the enrolment of cases, the sequencing and analysis received support from the NIHR BioResource – Rare Diseases. The NIHR BioResource is funded by the National Institute for Health Research (NIHR). We gratefully acknowledge the patients and their relatives who participated in the NIHR BioResource – Rare Diseases studies without whom much of this research would not be possible. The contributions by Jo Westmoreland (Medical Research Council {MRC} Laboratory of Molecular Biology) and by Dr Myrto Kostadima and Stuart Meacham (University of Cambridge, Ouwehand Group) in creating **Fig. 1**, **Fig. 2** and **Fig. 4** respectively are acknowledged.

C.L. is the recipient of a Clinical Research Training Fellowship award from the MRC and M.A.L. and C.L. are also supported by the Imperial College London NIHR Biomedical Research Centre. E.T. is supported by the NIHR BioResource and research in the Ouwehand laboratory receives support from the British Heart Foundation, European Commission, MRC, NHS Blood and Transplant, NIHR and Wellcome Trust.

Authorship

Contribution: C.L. wrote the paper; E.T. performed data analysis and edited the manuscript; M.A.L. and W.H.O. are Co-chairs of the BRIDGE-BPD consortium; K.F. and W.H.O. are Co-chair and Chair of the Scientific and Standardization Committee (SSC) on “Genomics in Thrombosis and Haemostasis” of the International Society on Thrombosis and Haemostasis. This SSC oversaw the development of the ThromboGenomics HTS test referred to in this review and described by Simeoni *et al* in this issue of *Blood*⁸². K.F., M.A.L. and W.H.O. edited the manuscript.

Conflict-of-interest disclosure

The authors declare no competing financial interests.

References

1. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature*. 2011;480(7376):201-208.
2. Westbury S, Turro E, Greene D, et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. *Genome Medicine*. 2015;7(1):36.
3. Dasouki MJ, Rafi SK, Olm-Shipman AJ, et al. Exome sequencing reveals a thrombopoietin ligand mutation in a Micronesian family with autosomal recessive aplastic anemia. *Blood*. 2013;122(20):3440-3449.
4. Dasouki M, Roberts J, Santiago A, Saadi I, Hovanes K. Confirmation and further delineation of the 3q26.33-3q27.2 microdeletion syndrome. *Eur J Med Genet*. 2014;57(2-3):76-80.
5. Ihara K, Ishii E, Eguchi M, et al. Identification of mutations in the c-mpl gene in congenital amegakaryocytic thrombocytopenia. *Proc Natl Acad Sci U S A*. 1999;96(6):3132-3136.
6. Savoia A, Dufour C, Locatelli F, et al. Congenital amegakaryocytic thrombocytopenia: clinical and biological consequences of five novel mutations. *Haematologica*. 2007;92(9):1186-1193.
7. Nichols KE, Crispino JD, Poncz M, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet*. 2000;24(3):266-270.
8. Yu C, Niakan KK, Matsushita M, Stamatoyannopoulos G, Orkin SH, Raskind WH. X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood*. 2002;100(6):2040-2045.
9. Song WJ, Sullivan MG, Legare RD, et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet*. 1999;23(2):166-175.
10. Hart A, Melet F, Grossfeld P, et al. Fli-1 is required for murine vascular and megakaryocytic development and is hemizygously deleted in patients with thrombocytopenia. *Immunity*. 2000;13(2):167-177.
11. Noetzel L, Lo RW, Lee-Sherick AB, et al. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet*. 2015;47(5):535-538.
12. Thompson AA, Nguyen LT. Amegakaryocytic thrombocytopenia and radio-ulnar synostosis are associated with HOXA11 mutation. *Nat Genet*. 2000;26(4):397-398.
13. Niihori T, Ouchi-Uchiyama M, Sasahara Y, et al. Mutations in MECOM, Encoding Oncoprotein EVI1, Cause Radioulnar Synostosis with Amegakaryocytic Thrombocytopenia. *Am J Hum Genet*. 2015;97(6):848-854.
14. Stevenson WS, Morel-Kopp MC, Chen Q, et al. GFI1B mutation causes a bleeding disorder with abnormal platelet function. *J Thromb Haemost*. 2013;11(11):2039-2047.
15. Monteferrario D, Bolar NA, Marneth AE, et al. A dominant-negative GFI1B mutation in the gray platelet syndrome. *N Engl J Med*. 2014;370(3):245-253.

16. Albers CA, Paul DS, Schulze H, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet.* 2012;44(4):435-439.
17. Pippucci T, Savoia A, Perrotta S, et al. Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. *Am J Hum Genet.* 2011;88(1):115-120.
18. Seri M, Cusano R, Gangarossa S, et al. Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Hegglin/Fechtner Syndrome Consortium. *Nat Genet.* 2000;26(1):103-105.
19. Kunishima S, Okuno Y, Yoshida K, et al. ACTN1 mutations cause congenital macrothrombocytopenia. *Am J Hum Genet.* 2013;92(3):431-438.
20. Nurden P, Debili N, Coupry I, et al. Thrombocytopenia resulting from mutations in filamin A can be expressed as an isolated syndrome. *Blood.* 2011;118(22):5928-5937.
21. Kunishima S, Kobayashi R, Itoh TJ, Hamaguchi M, Saito H. Mutation of the beta1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly. *Blood.* 2009;113(2):458-461.
22. Stritt S, Nurden P, Turro E, et al. A gain-of-function variant in DIAPH1 causes dominant macrothrombocytopenia and hearing loss. *Blood.* Prepublished Feb 24, 2016. <http://dx.doi.org/10.1182/blood-2015-10-675629>
23. Derry JM, Ochs HD, Francke U. Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. *Cell.* 1994;78(4):635-644.
24. Zhu Q, Zhang M, Blaese RM, et al. The Wiskott-Aldrich syndrome and X-linked congenital thrombocytopenia are caused by mutations of the same gene. *Blood.* 1995;86(10):3797-3804.
25. Albers CA, Cvejic A, Favier R, et al. Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. *Nat Genet.* 2011;43(8):735-737.
26. Kahr WH, Hinckley J, Li L, et al. Mutations in NBEAL2, encoding a BEACH protein, cause gray platelet syndrome. *Nat Genet.* 2011;43(8):738-740.
27. Gunay-Aygun M, Falik-Zaccai TC, Vilboux T, et al. NBEAL2 is mutated in gray platelet syndrome and is required for biogenesis of platelet alpha-granules. *Nat Genet.* 2011;43(8):732-734.
28. Castermans D, Volders K, Crepel A, et al. SCAMP5, NBEA and AMISYN: three candidate genes for autism involved in secretion of large dense-core vesicles. *Hum Mol Genet.* 2010;19(7):1368-1378.
29. Gissen P, Johnson CA, Morgan NV, et al. Mutations in VPS33B, encoding a regulator of SNARE-dependent membrane fusion, cause arthrogryposis-renal dysfunction-cholestasis (ARC) syndrome. *Nat Genet.* 2004;36(4):400-404.
30. Cullinane AR, Straatman-Iwanowska A, Zaucker A, et al. Mutations in VIPAR cause an arthrogryposis, renal dysfunction and cholestasis syndrome phenotype with defects in epithelial polarization. *Nat Genet.* 2010;42(4):303-312.
31. zur Stadt U, Rohr J, Seifert W, et al. Familial hemophagocytic lymphohistiocytosis type 5 (FHL-5) is caused by mutations in Munc18-2 and impaired binding to syntaxin 11. *Am J Hum Genet.* 2009;85(4):482-492.
32. Cote M, Menager MM, Burgess A, et al. Munc18-2 deficiency causes familial hemophagocytic lymphohistiocytosis type 5 and impairs cytotoxic granule exocytosis in patient NK cells. *J Clin Invest.* 2009;119(12):3765-3773.

33. Barbosa MD, Barrat FJ, Tchernev VT, et al. Identification of mutations in two major mRNA isoforms of the Chediak-Higashi syndrome gene in human and mouse. *Hum Mol Genet.* 1997;6(7):1091-1098.
34. Oh J, Bailin T, Fukai K, et al. Positional cloning of a gene for Hermansky-Pudlak syndrome, a disorder of cytoplasmic organelles. *Nat Genet.* 1996;14(3):300-306.
35. Dell'Angelica EC, Shotelersuk V, Aguilar RC, Gahl WA, Bonifacino JS. Altered trafficking of lysosomal proteins in Hermansky-Pudlak syndrome due to mutations in the beta 3A subunit of the AP-3 adaptor. *Mol Cell.* 1999;3(1):11-21.
36. Anikster Y, Huizing M, White J, et al. Mutation of a new gene causes a unique form of Hermansky-Pudlak syndrome in a genetic isolate of central Puerto Rico. *Nat Genet.* 2001;28(4):376-380.
37. Suzuki T, Li W, Zhang Q, et al. Hermansky-Pudlak syndrome is caused by mutations in HPS4, the human homolog of the mouse light-ear gene. *Nat Genet.* 2002;30(3):321-324.
38. Li W, Zhang Q, Oiso N, et al. Hermansky-Pudlak syndrome type 7 (HPS-7) results from mutant dysbindin, a member of the biogenesis of lysosome-related organelles complex 1 (BLOC-1). *Nat Genet.* 2003;35(1):84-89.
39. Zhang Q, Zhao B, Li W, et al. Ru2 and Ru encode mouse orthologs of the genes mutated in human Hermansky-Pudlak syndrome types 5 and 6. *Nat Genet.* 2003;33(2):145-153.
40. Morgan NV, Pasha S, Johnson CA, et al. A germline mutation in BLOC1S3/reduced pigmentation causes a novel variant of Hermansky-Pudlak syndrome (HPS8). *Am J Hum Genet.* 2006;78(1):160-166.
41. Cullinane AR, Curry JA, Carmona-Rivera C, et al. A BLOC-1 mutation screen reveals that PLDN is mutated in Hermansky-Pudlak Syndrome type 9. *Am J Hum Genet.* 2011;88(6):778-787.
42. Wright SD, Michaelides K, Johnson DJ, West NC, Tuddenham EG. Double heterozygosity for mutations in the platelet glycoprotein IX gene in three siblings with Bernard-Soulier syndrome. *Blood.* 1993;81(9):2339-2347.
43. Savoia A, Kunishima S, De Rocco D, et al. Spectrum of the mutations in Bernard-Soulier syndrome. *Hum Mutat.* 2014;35(9):1033-1045.
44. Newman PJ, Seligsohn U, Lyman S, Collier BS. The molecular genetic basis of Glanzmann thrombasthenia in the Iraqi-Jewish and Arab populations in Israel. *Proc Natl Acad Sci U S A.* 1991;88(8):3160-3164.
45. Nurden AT, Fiore M, Nurden P, Pillois X. Glanzmann thrombasthenia: a review of ITGA2B and ITGB3 defects with emphasis on variants, phenotypic variability, and mouse models. *Blood.* 2011;118(23):5996-6005.
46. Miller JL, Cunningham D, Lyle VA, Finch CN. Mutation in the gene encoding the alpha chain of platelet glycoprotein Ib in platelet-type von Willebrand disease. *Proc Natl Acad Sci U S A.* 1991;88(11):4761-4765.
47. Randi AM, Rabinowitz I, Mancuso DJ, Mannucci PM, Sadler JE. Molecular basis of von Willebrand disease type IIB. Candidate mutations cluster in one disulfide loop between proposed platelet glycoprotein Ib binding sequences. *J Clin Invest.* 1991;87(4):1220-1226.

48. Turro E, Greene D, Wijgaerts A, et al. A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies. *Sci Transl Med*. 2016;8(328):328ra330.
49. Morison IM, Cramer Borde EM, Cheesman EJ, et al. A mutation of human cytochrome c enhances the intrinsic apoptotic pathway but causes only thrombocytopenia. *Nature Genetics*. 2008;40(4):387-389.
50. Misceo D, Holmgren A, Louch WE, et al. A dominant STIM1 mutation causes Stormorken syndrome. *Hum Mutat*. 2014;35(5):556-564.
51. Nesin V, Wiley G, Kousi M, et al. Activating mutations in STIM1 and ORAI1 cause overlapping syndromes of tubular myopathy and congenital miosis. *Proc Natl Acad Sci U S A*. 2014;111(11):4197-4202.
52. Lacruz RS, Feske S. Diseases caused by mutations in ORAI1 and STIM1. *Ann N Y Acad Sci*. 2015;1356:45-79.
53. Fletcher SJ, Johnson B, Lowe GC, et al. SLFN14 mutations underlie thrombocytopenia with excessive bleeding and platelet secretion defects. *J Clin Invest*. 2015;125(9):3600-3605.
54. Marconi C, Di Buduo CA, Barozzi S, et al. SLFN14-related thrombocytopenia: identification within a large series of patients with inherited thrombocytopenia. *Thromb Haemost*. 2016;115(5).
55. Paterson AD, Rommens JM, Bharaj B, et al. Persons with Quebec platelet disorder have a tandem duplication of PLAU, the urokinase plasminogen activator gene. *Blood*. 2010;115(6):1264-1266.56. Dumont B, Lasne D, Rothschild C, et al. Absence of collagen-induced platelet activation caused by compound heterozygous GPVI mutations. *Blood*. 2009;114(9):1900-1903.
57. Fang J, Nurden P, North P, et al. C560Rbeta3 caused platelet integrin alphaIIb beta3 to bind fibrinogen continuously, but resulted in a severe bleeding syndrome and increased murine mortality. *J Thromb Haemost*. 2013;11(6):1163-1171.
58. Ghevaert C, Salsmann A, Watkins NA, et al. A nonsynonymous SNP in the ITGB3 gene disrupts the conserved membrane-proximal cytoplasmic salt bridge in the alphaIIb beta3 integrin and cosegregates dominantly with abnormal proplatelet formation and macrothrombocytopenia. *Blood*. 2008;111(7):3407-3414.
59. Svensson L, Howarth K, McDowall A, et al. Leukocyte adhesion deficiency-III is caused by mutations in KINDLIN3 affecting integrin activation. *Nat Med*. 2009;15(3):306-312.
60. Canault M, Ghalloussi D, Grosdidier C, et al. Human CalDAG-GEFI gene (RASGRP2) mutation affects platelet function and causes severe bleeding. *J Exp Med*. 2014;211(7):1349-1362.
61. Hollopeter G, Jantzen HM, Vincent D, et al. Identification of the platelet ADP receptor targeted by antithrombotic drugs. *Nature*. 2001;409(6817):202-207.
62. Hirata T, Kakizuka A, Ushikubi F, Fuse I, Okuma M, Narumiya S. Arg60 to Leu mutation of the human thromboxane A2 receptor in a dominantly inherited bleeding disorder. *J Clin Invest*. 1994;94(4):1662-1667.
63. Genevieve D, Proulle V, Isidor B, et al. Thromboxane synthase mutations in an increased bone density disorder (Ghosal syndrome). *Nat Genet*. 2008;40(3):284-286.

64. Adler DH, Cogan JD, Phillips JA, 3rd, et al. Inherited human cPLA(2 α) deficiency is associated with impaired eicosanoid biosynthesis, small intestinal ulceration, and platelet dysfunction. *J Clin Invest*. 2008;118(6):2121-2131.
65. Suzuki J, Umeda M, Sims PJ, Nagata S. Calcium-dependent phospholipid scrambling by TMEM16F. *Nature*. 2010;468(7325):834-838.
66. Gresele P. Diagnosis of inherited platelet function disorders: guidance from the SSC of the ISTH. *J Thromb Haemost*. 2015;13(2):314-322.
67. Bolton-Maggs PH, Chalmers EA, Collins PW, et al. A review of inherited platelet disorders with guidelines for their management on behalf of the UKHCDO. *Br J Haematol*. 2006;135(5):603-633.
68. Zhang MY, Churpek JE, Keel SB, et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet*. 2015;47(2):180-185.
69. Noris P, Perrotta S, Seri M, et al. Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families. *Blood*. 2011;117(24):6673-6680.
70. Shin CR, Kim MO, Li D, et al. Outcomes following hematopoietic cell transplantation for Wiskott-Aldrich syndrome. *Bone Marrow Transplant*. 2012;47(11):1428-1435.
71. Hacein-Bey Abina S, Gaspar HB, Blondeau J, et al. Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome. *JAMA*. 2015;313(15):1550-1563.
72. Ballmaier M, Germeshausen M. Congenital amegakaryocytic thrombocytopenia: clinical presentation, diagnosis, and treatment. *Semin Thromb Hemost*. 2011;37(6):673-681.
73. Gresele P, Harrison P, Bury L, et al. Diagnosis of suspected inherited platelet function disorders: results of a worldwide survey. *J Thromb Haemost*. 2014;12(9):1562-1569.
74. Quiroga T, Goycoolea M, Panes O, et al. High prevalence of bleeders of unknown cause among patients with inherited mucocutaneous bleeding. A prospective study of 280 patients and 299 controls. *Haematologica*. 2007;92(3):357-365.
75. Cattaneo M. The platelet P2Y₁(2) receptor for adenosine diphosphate: congenital and drug-induced defects. *Blood*. 2011;117(7):2102-2112.
76. Matus V, Valenzuela G, Saez CG, et al. An adenine insertion in exon 6 of human GP6 generates a truncated protein associated with a bleeding disorder in four Chilean families. *J Thromb Haemost*. 2013;11(9):1751-1759.
77. Hermans C, Wittevrongel C, Thys C, Smethurst PA, Van Geet C, Freson K. A compound heterozygous mutation in glycoprotein VI in a patient with a bleeding disorder. *J Thromb Haemost*. 2009;7(8):1356-1363.
78. Mumford AD, Nisar S, Darnige L, et al. Platelet dysfunction associated with the novel Trp29Cys thromboxane A₂ receptor variant. *J Thromb Haemost*. 2013;11(3):547-554.
79. Zwaal RFA, Comfurius P, Bevers EM. Scott syndrome, a bleeding disorder caused by defective scrambling of membrane phospholipids. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. 2004;1636(2-3):119-128.

80. Bottega R, Marconi C, Faleschini M, et al. ACTN1-related thrombocytopenia: identification of novel families for phenotypic characterization. *Blood*. 2015;125(5):869-872.
81. Noris P, Favier R, Alessi MC, et al. ANKRD26-related thrombocytopenia and myeloid malignancies. *Blood*. 2013;122(11):1987-1989.
82. Simeoni I, Stephens JC, Hu F et al. A comprehensive high-throughput sequencing test for the diagnosis of inherited bleeding, thrombotic and platelet disorders. *Blood*. 2016;In Press.
83. The 100,000 Genomes Project. <http://www.genomicsengland.co.uk/>.
84. Greene Daniel NB, Richardson Sylvia, Turro Ernest. Phenotype similarity regression for identifying the genetic determinants of rare diseases. *American Journal of Human Genetics*. 2016.
85. Mumford AD, Frelinger Iii AL, Gachet C, et al. A review of platelet secretion assays for the diagnosis of inherited platelet secretion disorders. *Thrombosis and Haemostasis*. 2015;114(7):14-25.
86. Rao AK. Inherited platelet function disorders: overview and disorders of granules, secretion, and signal transduction. *Hematol Oncol Clin North Am*. 2013;27(3):585-611.
87. Watson SP, Lowe GC, Lordkipanidzé M, Morgan NV, The Gc. Genotyping and phenotyping of platelet function disorders. *Journal of Thrombosis and Haemostasis*. 2013;11:351-363.
88. Leo VC, Morgan NV, Bem D, et al. Use of next-generation sequencing and candidate gene analysis to identify underlying defects in patients with inherited platelet function disorders. *J Thromb Haemost*. 2015;13(4):643-650.
89. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
90. Walter K, Min JL, Huang J, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526(7571):82-90.
91. Exome Aggregation Consortium, Lek M, Karczewski K et al. Analysis of protein-coding genetic variation in 60,706 humans. Prepublished October 30 2015. BioRxiv. <http://dx.doi.org/10.1101/030338>
92. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
93. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;508(7497):469-476.
94. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862-868.
95. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*. 2009;84(4):524-533.
96. Exome Variant Server <http://evs.gs.washington.edu/EVS/> Accessed March 10th 2016.
97. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and

- molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet.* 2014;133(1):1-9.
98. Xue Y, Chen Y, Ayub Q, et al. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet.* 2012;91(6):1022-1032.
99. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. *Cell.* 2011;144(6):986-998.
100. Rolland T, Taşan M, Charleatoux B, et al. A Proteome-Scale Map of the Human Interactome Network. *Cell*;159(5):1212-1226.
101. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57-74.
102. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317-330.
103. Chen L, Kostadima M, Martens JH, et al. Transcriptional diversity during lineage commitment of human blood progenitors. *Science.* 2014;345(6204):1251033.
104. Kohler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 2014;42(Database issue):D966-974.
105. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature.* 2015;519(7542):223-228.
106. Masino AJ, Dechene ET, Dulik MC, et al. Clinical phenotype-based gene prioritization: an initial study using semantic similarity and the human phenotype ontology. *BMC Bioinformatics.* 2014;15:248.
107. Zemojtel T, Kohler S, Mackenroth L, et al. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci Transl Med.* 2014;6(252):252ra123.
108. Bluteau D, Balduini A, Balayn N, et al. Thrombocytopenia-associated mutations in the ANKRD26 regulatory region induce MAPK hyperactivation. *J Clin Invest.* 2014;124(2):580-591.
109. Londin E, Loher P, Telonis AG, et al. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc Natl Acad Sci U S A.* 2015;112(10):E1106-1115.
110. Tijssen MR, Cvejic A, Joshi A, et al. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell.* 2011;20(5):597-609.
111. Vasquez LJ, Mann AL, Chen L, Soranzo N. From GWAS to function: lessons from blood cells. *ISBT Science Series.* 2016;11:211-219.
112. Freson K, De Vos R, Wittevrongel C, et al. The TUBB1 Q43P functional polymorphism reduces the risk of cardiovascular disease in men by modulating platelet function and structure. *Blood.* 2005;106(7):2356-2362.
113. Auer PL, Teumer A, Schick U, et al. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nat Genet.* 2014;46(6):629-634.
114. Kunishima S, Nishimura S, Suzuki H, Imaizumi M, Saito H. TUBB1 mutation disrupting microtubule assembly impairs proplatelet formation and results in congenital macrothrombocytopenia. *Eur J Haematol.* 2014;92(4):276-282.

115. Jupe S, Akkerman JW, Soranzo N, Ouwehand WH. Reactome - a curated knowledgebase of biological pathways: megakaryocytes and platelets. *J Thromb Haemost.* 2012;10(11):2399-2402.
116. Mumford AD, Dawood BB, Daly ME, et al. A novel thromboxane A2 receptor D304N variant that abrogates ligand binding in a patient with a bleeding diathesis. *Blood.* 2010;115(2):363-369.
117. Kamae T, Kiyomizu K, Nakazawa T, et al. Bleeding tendency and impaired platelet function in a patient carrying a heterozygous mutation in the thromboxane A2 receptor. *J Thromb Haemost.* 2011;9(5):1040-1048.
118. Nisar SP, Lordkipanidze M, Jones ML, et al. A novel thromboxane A2 receptor N42S variant results in reduced surface expression and platelet dysfunction. *Thromb Haemost.* 2014;111(5):923-932.
119. Freson K, Thys C, van Geet C. First report of transheterozygosity in a patient with an inherited platelet bleeding disorder due to mutations in the Thromboxane A2 receptor (TBXA2R) and cyclooxygenase1 (PTGS1). American Society of Hematology Meeting; 2011.
120. Hu Y, Yan C, Hsu CH, et al. OmicCircos: A Simple-to-Use R Package for the Circular Visualization of Multidimensional Omics Data. *Cancer Inform.* 2014;13:13-20.

Figure Legends

Figure 1: The 51 genes underlying Inherited Platelet Disorders.

The cartoon depicts the process of megakaryopoiesis and platelet formation. Each of the 51 known inherited platelet disorder (IPD) genes are indicated and categorised according to their effect on megakaryocyte and platelet biology. IPDs typically associated with phenotypes outside of the blood system are indicated by *; HSC, hematopoietic stem cell.

Figure 2: Expression levels of 51 genes underlying Inherited Platelet Disorders across hematopoietic stem and progenitor cells.

High relative expression is shown in red and low relative expression in blue. The expression of each gene is normalized compared to a mean expression of zero across all the samples. Genes are ordered and colour-coded according to their predicted effect on platelet biology (as in **Fig. 1**). Information about the levels of transcripts for the 51 genes determined by RNA-seq was retrieved from Chen *et al*¹⁰³. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythrocyte precursor; EB, erythroblast; MK, megakaryocyte.

Figure 3: Genomic location of the 51 genes underlying Inherited Platelet Disorders.

Circos diagram¹²⁰ illustrating the location of known IPD genes across human chromosomes. **Track 1:** Cytoband with chromosome name with centromeres in blue. **Track 2:** Genomic location of 51 established inherited platelet disorder (IPD) genes and the year in which variants in the gene were first identified as a cause of IPD in humans in brackets. Gene names in red represent genes identified by high throughput sequencing. **Track 3:** Log₁₀ of the number of amino acids encoded by the reference CCDS transcript. Log₁₀ scale is indicated at 12 o'clock. **Track 4:** Log₁₀ of the number of rare variants predicted to affect amino acid sequence observed in 6,390 individuals enrolled to the NIHR BioResource – Rare Diseases; Log₁₀ scale is indicated at 12 o'clock.

Figure 4: Protein-protein interaction network reflecting the molecules and pathways implicated in megakaryopoiesis, the formation of platelets, thrombosis and hemostasis.

A. Protein-protein interaction network (PPIN) of 1,684 nodes (proteins) connected by 5,360 edges (biochemical reactions). The 1,517 first-order interacting nodes and all but 24 of the 5,360 edges were obtained from the Reactome (n=3,625) and IntAct (n=1,711) databases. The 24 edges were added on basis of manual literature curation. The 200 baits are colored as per the Venn diagram **(B)** except for the eight baits present in more than one category, which are pink and the eight prototype proteins involved in the synthesis of thromboxane and signalling via the thromboxane-thromboxane receptor (Tbxa2r) pathway. The Venn diagram shows the three gene sets in ochre, blue and purple for the ThromboGenomics HTS test platform gene set⁸², the platelet volume and count GWAS gene set¹ or the gene set identified by ChIP-seq in human megakaryocytes and showing binding of all five

transcription factors (Fli1, Gata1, Gata2, Runx1, Tal1) at their promoter¹¹⁰, respectively. **C and D.** Sub-networks retrieved from the PPIN in **(A)**. **C.** A sub-network of 156 nodes and 874 edges obtained by retrieving the first order interactors of Tbx2r, the receptor for thromboxane. **D.** A sub-network of 26 nodes and 42 edges involved in the synthesis of thromboxane and obtained by selecting the first order interactors of Tbxas1 (Thromboxane synthase 1) and Pla2g4a (Phospholipase A2). The red nodes in **C** and **D** are a set of prototype proteins related with thromboxane synthesis and signalling and the other colored nodes are baits. The surface area of the red colored nodes in **A** and all colored nodes in **C** and **D** equate with their transcript level determined by sequencing of RNA from human megakaryocytes (data retrieved from Chen *et al*)¹⁰³. An interactive version of the network, containing gene expression levels and other annotation features is available for download in Cytoscape format from Supplemental data.

Fig. 1

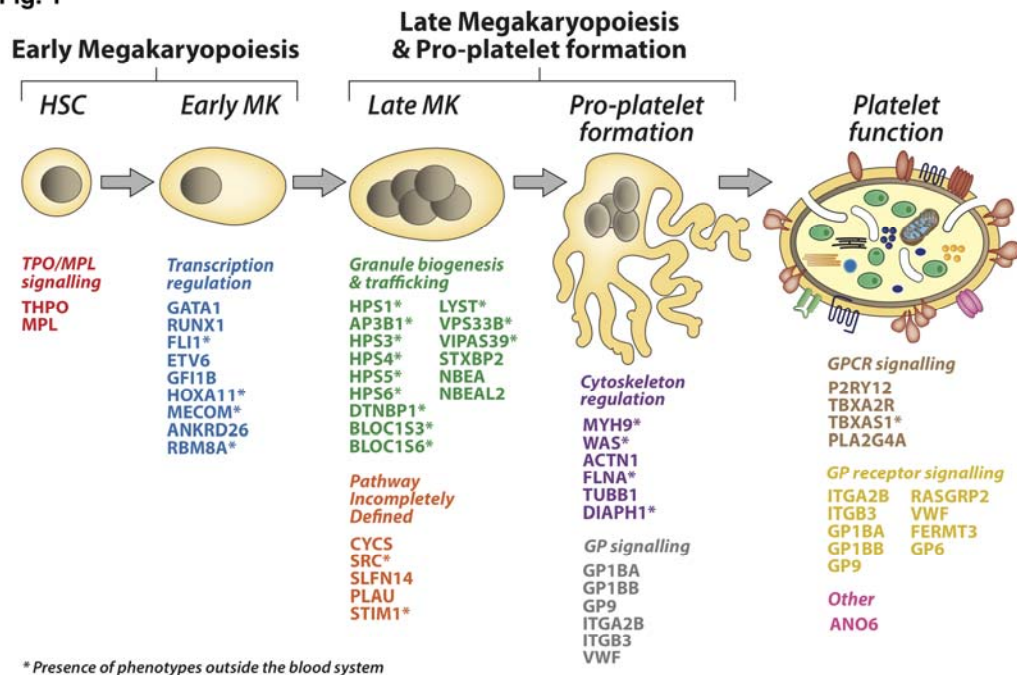


Fig. 2

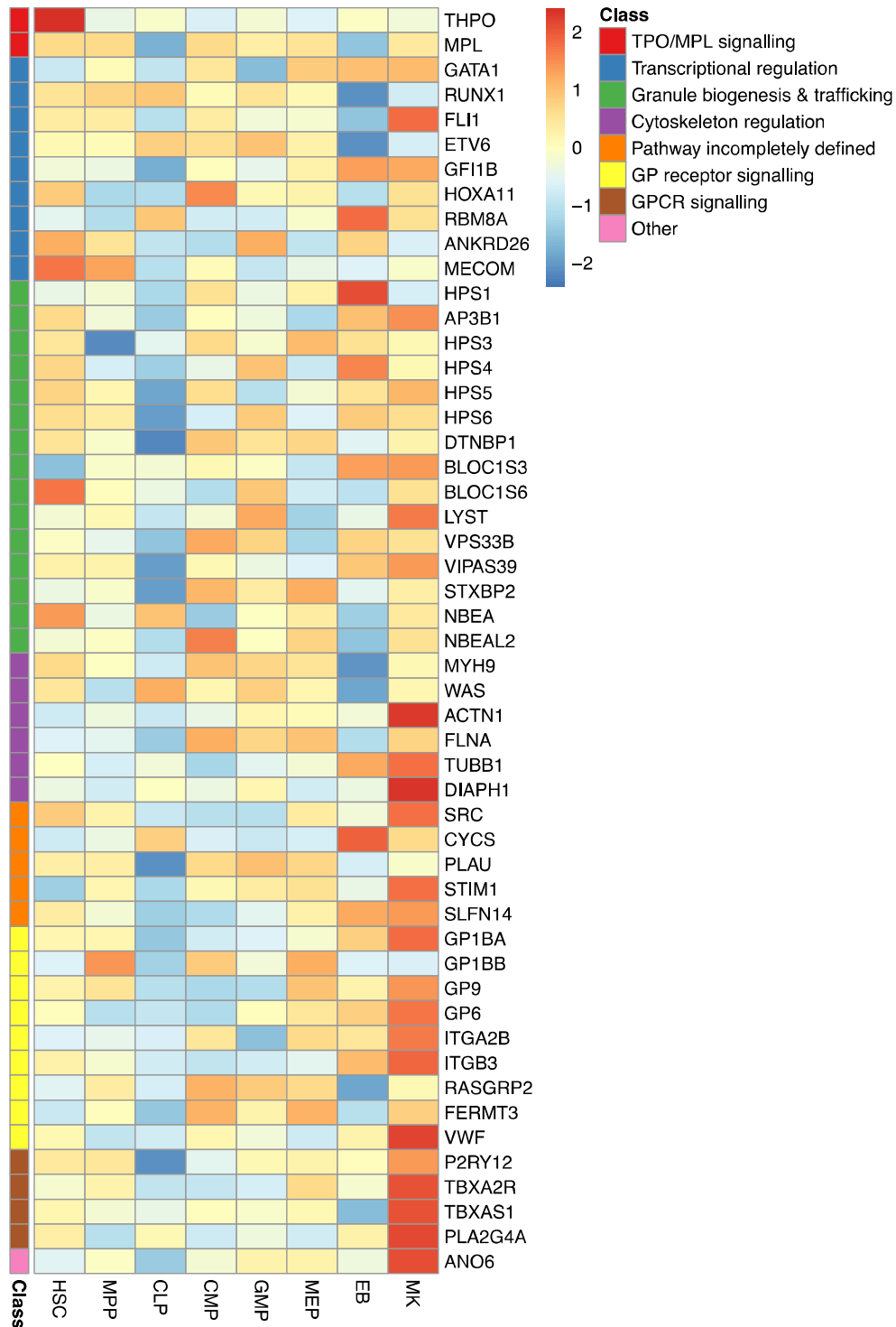
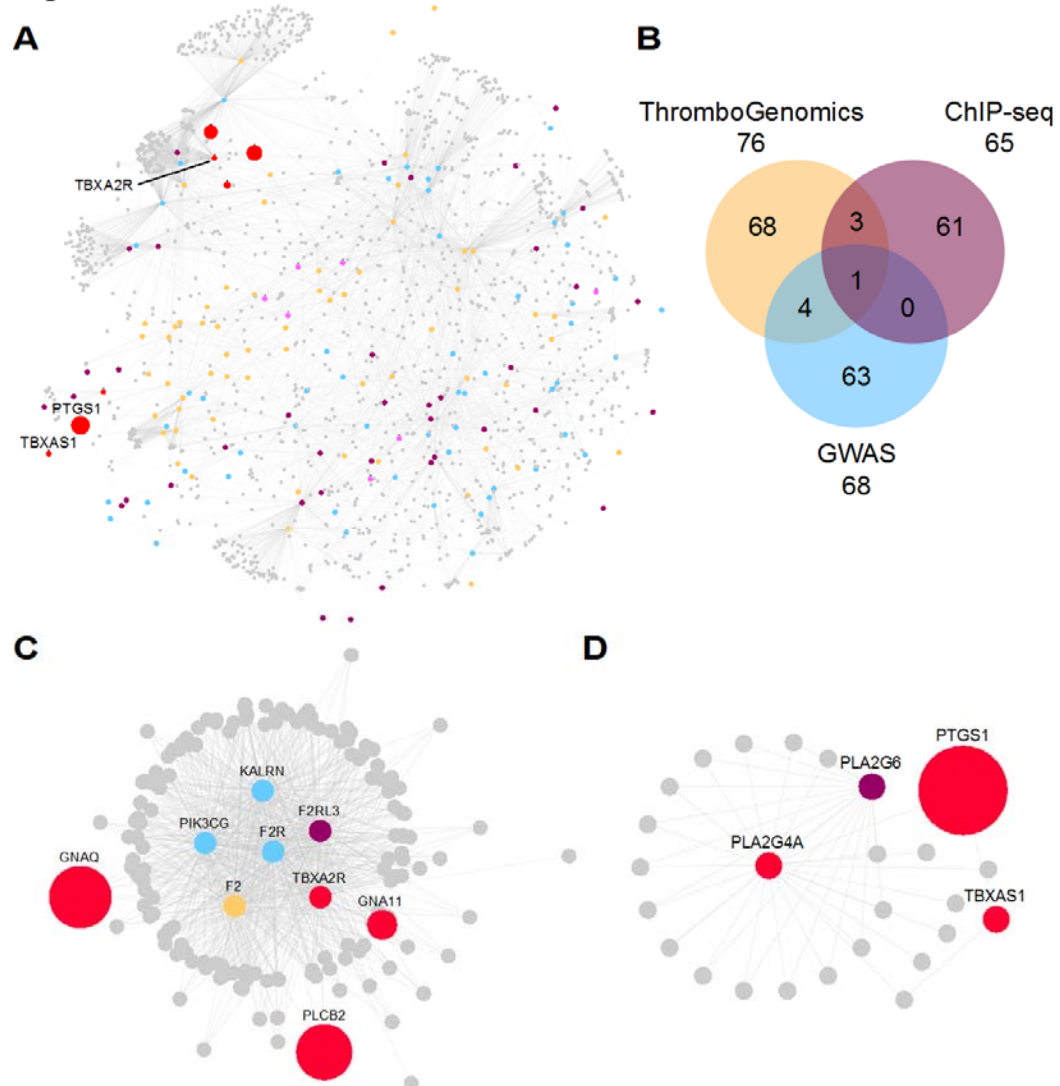


Fig. 4



Glossary of Terms

Inherited Platelet Disorder (IPD): Any genetic abnormality of platelet number, morphology or function in humans

IPD Gene: A gene in which a DNA variant has been shown to cause an IPD in humans

Bleeding, Thrombotic and Platelet Disorders (BPD): Term used by the NIHR BioResource and the ThromboGenomics consortium to describe cases with genetic variants leading to a tendency to bleed abnormally, or a tendency to abnormal thrombosis or an abnormality of the count, volume, morphology or function of platelets.

High-Throughput Sequencing (HTS): Term encompassing a variety of modern techniques, which allow the sequencing of DNA and RNA more quickly and cheaply than traditional Sanger sequencing (*also referred to as next-generation sequencing*).

Minor Allele Frequency (MAF): The frequency at which the least common allele at a particular variant position occurs in a population.

Whole Exome Sequencing (WES): The sequencing by HTS of the coding portion of the genome comprising ~65 million bases.

Whole Genome Sequencing (WGS): The sequencing by HTS of the entire genome comprising ~3.2 billion bases.

Single Nucleotide Variant (SNV): A single base pair alteration in a DNA sequence compared to the reference genome for a particular species. A SNV does not necessarily alter the function of a gene. SNVs can be common (MAF>5%), low frequency (MAF between 1% and 5%) or rare (MAF<1%).

Loss of Function (LOF): A variant allele causing the complete or partial absence of normal gene function compared to the wild-type allele.

Gain of Function (GOF): A variant allele causing a gain of normal gene function compared to the wild-type allele.

5-prime untranslated region (5'UTR): The section of the message RNA (mRNA) upstream from the start codon (where translation of mRNA starts). The 5'UTR is important in regulating translation.

ChIP-seq (Chromatin immunoprecipitation combined with HTS): A powerful method for identifying modification of DNA-associated proteins like histones or used for mapping the binding sites of proteins like transcription factors; the method is used extensively to enhance the functional annotation of the non-coding portion of the genome.

Genome Wide Association Study (GWAS): A study comparing the frequencies of millions of SNVs across thousands of people with and without a particular disease or assessing the effect of these SNVs on a quantitative trait. This type of study explores the genome in a hypothesis-free approach for variants that confer a risk for a particular disease or are exerting an effect on the value of a certain trait.

Protein-protein interaction network (PPIN): A representation of physical contacts established between proteins. PPINs can encompass knowledge on biochemical pathways derived from thousands of experiments and may be represented graphically in a network diagram.



blood®

Prepublished online April 19, 2016;
doi:10.1182/blood-2016-03-378588

Inherited platelet disorders: towards DNA-based diagnosis

Claire Lentaigne, Kathleen Freson, Michael A. Laffan, Ernest Turro and Willem H. Ouwehand

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
<http://www.bloodjournal.org/site/misc/rights.xhtml#reprints>

Information about subscriptions and ASH membership may be found online at:
<http://www.bloodjournal.org/site/subscriptions/index.xhtml>

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include digital object identifier (DOIs) and date of initial publication.